UNIT 11 POISSON QUEUES

Structure **Page No**

11.1 INTRODUCTION

Queues (Waiting Lines) are a part of everyday life. We all wait in queues to buy a movie ticket, make bank deposit, pay for groceries, mail a package, obtain food in cafeteria, and start a ride in an amusement park, etc. We have become habitual to certain amount of waiting, but still get irritated by usually long waits.

So we can say waiting line arises when the demand for a service facility exceeds the capacity of that facility, i.e. the customers do not get service immediately upon request but must wait, or the service facility stands idle and waits for customers.

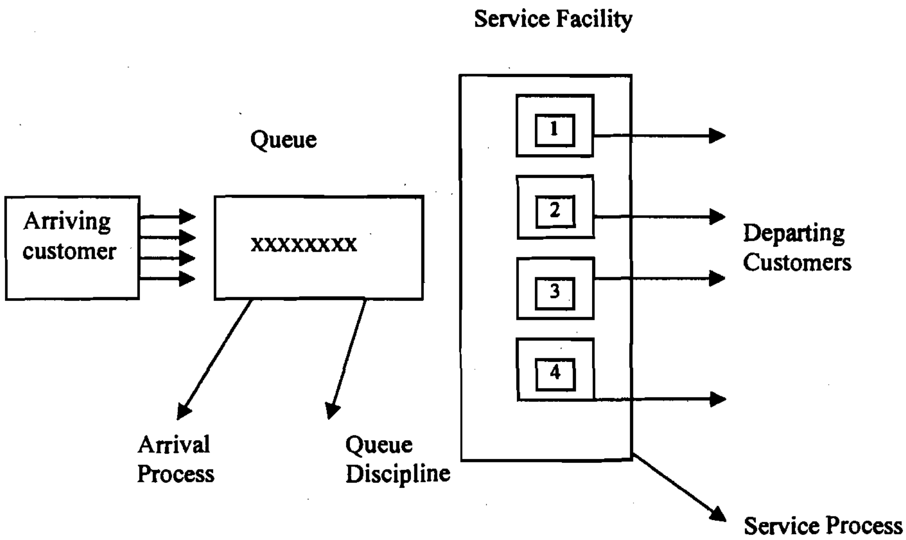
However, having to wait is not just a pretty personal annoyance. The amount of time that we waste by waiting in queues is a major factor in both the quality of life and the efficiency of the nation's economy. Waiting lines or queues are a common occurrence both in everyday life and in variety of business and industrial situations. Most waiting line problems are centered about the question of finding the ideal level of services that a firm should provide for a high level of satisfaction to the customer at a low cost to the firm. The applications of queueing can be seen in the following situations.

- to decide the number of cash register check out positions in super market.
- to decide how many pumps should be opened and how many attendants should be on duty in Gasoline stations.
- to determine the optimal number of mechanics to have on duty in each shift to repair machines that break down in manufacturing plants.
- to decide the number of teller windows to keep open to serve customers during various hours of the day in Banks.

Queueing Theory had its beginning in the research work of a Danish engineer named A. K. Erlang. In 1909, Erlang experimented with fluctuating demand in telephone traffic. Eight years later he published a report addressing the delays in automatic dialing equipment. At the end of World War II, Erlang's early work was extended to more general problems and to business applications of waiting lines.

Queueing theory is the study of waiting lines in all these various daytoday life problems. It uses queueing models to represent the various types of queueing systems (system that involve queues of some kind) that arise in practice. Formulas derived for each model results in how the corresponding queueing system should perform, including the average amount of waiting that will occur, under a variety of circumstances.

Therefore, these queueing models are very helpful for determining the performance indices of the situation modeled.



Schematic Representation of a Queueing Problem

This diagram is a systematic presentation of a queueing system. It gives a better understanding of discussion given above. In this unit, we shall discuss two simple queueing models $M/M/1/\infty$ and $M/M/k/\infty$.

Objectives

After studying this unit, you should be able to:

- identify the occurrence of queueing in real life situations;
- describe the characteristics of a queueing problem;
- use the statistical methods necessary to analyze queueing problem;
- apply the two common queueing models in suitable problems;
- estimate the optimum parameters of a queueing model.

In the next section, we will define the various basic components of a queueing model by a technical and systematic approach.

11.2 CHARACTERISTICS OF QUEUEING MODELS

Let us discuss the basic features that characterize the queueing models.

a) Input (or Arrival Pattern)

The input describes the way in which the customers arrive and join the system for service. Generally, the customers arrive in a random fashion, which is not worth making the prediction. In that case, the arrival pattern can best be described in terms of probabilities. Consequently, the probability distribution for inter arrival times (the time between the two successive arrivals) or the probability distribution of number of customer arriving in time t must be defined. Two patterns that can be observed in practice are:

i) Arrivals are of regular intervals (like a candidate for interview for 45 minutes)

ii) There is general probability distribution (perhaps exponential) of time between successive arrivals.

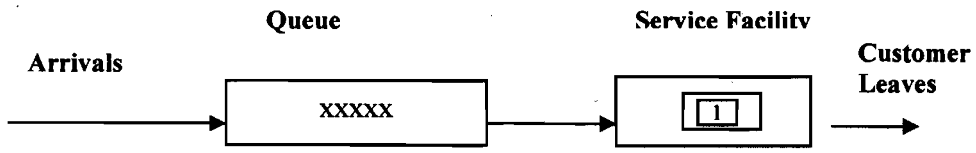
b) The Service Mechanism

It is specified when it is known how many customers can be served at a time, what the statistical distribution of service time is, and when the service is available. It is true in most situations that service time is a random variable with the same distribution for all arrivals, but cases also occur where there are two or more classes of customers (e.g. different types of machines waiting repair) each with a different service time distribution.

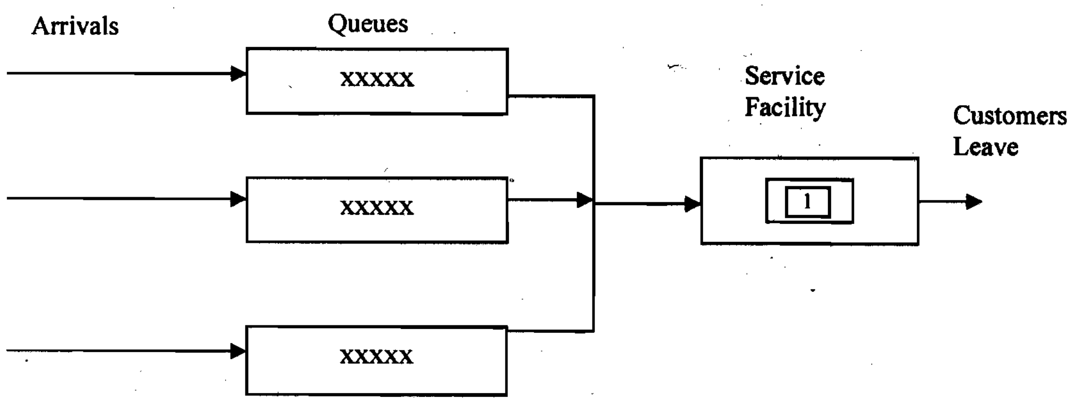
c) Servers

The service may be offered through a single server such as ticket counter or through several channels such as trains arriving in a station with several platforms. The time required for servicing, a unit (or a group in case of batch service) is called the **service time.** Service systems can be classified in terms of their number of service channels, or the number of servers.

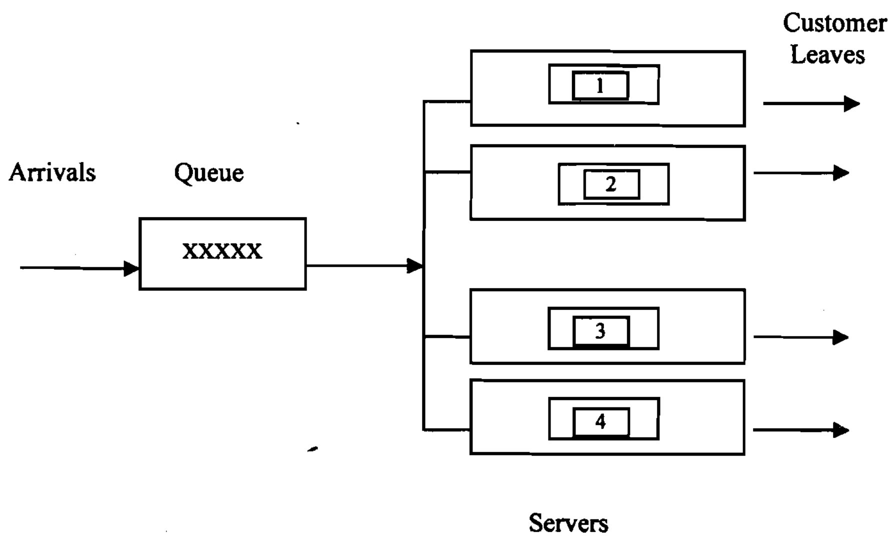
i) **Single Server-Single Queue:** The models that involve one queue-one service station facility are called single server models where customer waits till the service point is ready to take him for servicing. Patients arriving at a clinic of a doctor is an example of a single server facility.



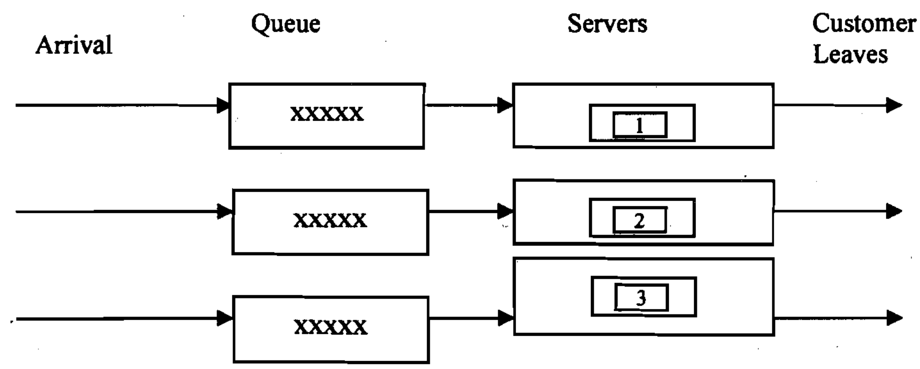
ii) **Single Server-Several Queues:** In this type of facility, there are several queues and the customer may join any one of these but there is only one service channel. For example one cash counter is an electricity office where the customers in two queues (for males and females) can make payments in respect of their electricity bills.



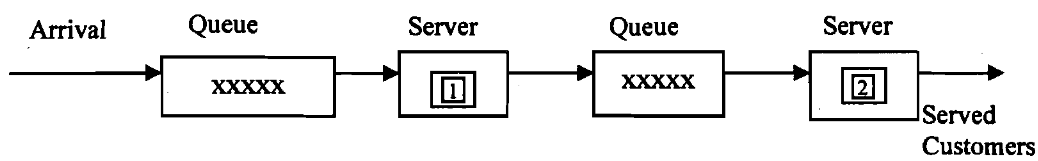
iii) **Several (Parallel) Servers-Single Queue:** In this type of model, there is more than one server and each server provides the same type of service. The customers wait in a single queue until one of the service channels is ready to take them in for service.



iv) **Several Servers-Several Queues:** This type of model consists of several servers where each of the servers has a different queue. Different railway ticket booking windows in a railway station where the customers can book their tickets in respect of their reservations at any window provides an example of this type of model.



v) **Service Facility in series:** In this customer enters the system, and is served in many phases. First he is served at one server and after getting one phase of service at that server it moves ahead to get service from the next server and so on. The service is completed after the final phase of service from the last server in the series.



d) Queue Discipline

The queue discipline is the rule determining the formation of the queue, the manner of the customer's behavior while waiting, and the manner in which they are chosen for service. Properties of queueing system which are concerned with waiting times, in general, depend on queue discipline. The following notations are used for describing the nature of service discipline:

i) **First Come, First Served (FCFS) or First in First Out (FIFO)**

This is a simplest and most commonly used queue discipline. According to this discipline, the customers are served in the order of their arrival. This service discipline may be seen at a cinema ticket window, at a railway ticket window, etc.

ii) **Last Come, First Served (LCFS) or Last in First Out (LIFO)**

This discipline may be seen in big warehouses where the units (items) which come and are stored last are taken out (served) first.

iii) **Service in Random Order (SIRO)**

In this Queueing discipline, the customers are selected randomly for service. In this, every customer in the queue is equally likely to be selected. The time of arrival of the customers is, therefore, of no relevance in such a case.

iv) **Pre-emptive priority**

In this queuing discipline, the service of a customer having lower priority is interrupted and the service of customer having highest priority is started. After the highest priority custom is served, the lower priority customer is resumed.

e) **Customer's Behaviour**

Another thing which needs to be considered in the queueing structure is the behaviour or attitude of the customers entering the queueing system. A customer who waits for service irrespective of the long waiting time is called a patient customer. Whereas a customer, who waits for a certain amount of time in the queue and leaves the service system without getting service due to certain reasons such as a long queue in front of him is called an impatient customer.

Now let's check out the customer behaviour in the queueing system:

i) **Balking:** A customer may leave the queue even before joining the queue because the queue is too long and he has no time to wait, or there is not sufficient waiting space. This is called balking.

ii) **Reneging:** Customers after joining the queue wait for sometime and leave the service system due to intolerable delay, so they renege.

iii) **Priorities:** In certain applications some customers are served before others regardless of their order of arrival. These customers have priority of service over others.

iv) **Jockeying:** Where there are many identical servers with each server having a different queue a customer may jockey from one waiting line to another in hope of getting served quickly.

f) **Size of Population**

The collection of potential customers may be very large or of a moderate size. In a railway booking counter the total number of potential passengers is so large that although finite theoretically it can be regarded infinity for all practical purposes. The assumption of infinite population is very convenient for a queueing model.

g) **Length of the Queue**

Sometimes because of capacity or space limitations only a finite number of customers are allowed to stay in the system although the arrival stream is infinite. The queueing process specifies to the number of queues, and their respective lengths. The number of queues depends upon the layout and service offered.

E1) Give one real life situation of each of the terms given above.

So these are some of the important components of the queueing system. Now we are going to study various waiting line models to evaluate the size of queue, time to wait in a queue and various other factors. But before moving to models, first we need to establish some statistical results for the models.

11.3 STATISTICAL DISTRIBUTIONS USED IN QUEUEING THEORY

In this section, let us discuss various statistical distributions which could be used for queueing theory analysis purposes.

i) Degenerate (or Deterministic) Distribution

The random variable in this distribution takes a constant value with probability one. This could be used, for example, to model the input of parts to a manufacturing machine, if one knew that the input of a part occurred at every, say c, units of time.

ii) Markovian or the Negative Exponential and Erlang Distributions

Erlang distribution is named after A.K. Erlang a denish engineer (1878-1929), one of the fathers of queueing theory.

The negative exponential distribution is the widely used distribution in queueing theory to model the probability law of interarrival time or the service time. It's p.d.f. (Probability density function) is:

$$f(x) = \frac{1}{\alpha} e^{-\frac{x}{\alpha}} \quad 0 \le x < \infty,\ \alpha > 0.$$

As it is clear from the above, the exponential distribution has one parameter which is denoted by the letter α and is the mean of this distribution. The exponential distribution is a special case of the Erlang distribution. The p.d.f. of the Erlang distribution is:

$$f(x) = \frac{b^{k+1} x^k e^{-bx}}{k!} \quad 0 \le x < \infty,\ b > 0,\ k = 0, 1, 2, \ldots$$

It is the distribution of sum of k independent exponential random variables each with parameter $\frac{1}{b}$.

iii) General Distribution

This does not refer to any particular statistical distribution, rather, an arbitrary one. In other words, every distribution is a special case of the general distribution, and all results derived for queueing system from use of the general distribution can be applied to any other distribution, incorporating the appropriate restrictions of special case. For queueing theory purposes, one normally requires the mean and variance of the distribution only in case General Distribution is assumed in the model.

Now we must be aware of the different queueing models and these models can be easily described and understood by the use of the Kendall's notation. This notation is discussed in the following section.

Notations for queueing systems are proposed by D.G. Kendall.

11.4 KENDALL'S NOTATION

There is a standard notation for classifying queueing systems into different types, which is known as Kendall's notation. The idea of this notation is to specify a particular queueing system by a six terms code separated by slashes. Thus, any system is described by a notation:

A / B / C / D / E / F

where the symbols A, B, C, etc. stand for:

A	Distribution of interarrival times of customers
B	Distribution of service times
C	Number of servers
D	Maximum number of customers which can be accommodated in system at any time
E	Calling population size
F	Queue discipline such as FCFS or LCFS

A and **B** can take any of following distribution types:

M	Exponential Distribution (Markovian)
D	Degenerate (or Deterministic) Distribution
E_k	Erlang Distribution (k = shape parameter)
GI	General independent inter arrival time distribution
G	General Distribution (arbitrary distribution)

C, **D** and **E** may either variables or numbers. To describe a queueing system **A**, **B** and C are always given, however **E**, **F** or **D**, **E**, **F** may be omitted if we intend to assume their usual values of infinite capacity infinite source population and FCFS discipline.

Example 1: Consider the following queues:

i) $D/M/n$: It represents a queueing system with a degenerate distribution for the interarrival times of customers, an exponential distribution for service times of customers for each of n servers.

ii) $E_k/E_1/1$: This means a queueing system with an Erlang distribution for the interarrival times of customers (with a shape parameter of k), an exponential distribution for service times of customers (with a shape parameter of 1), and a single server.

iii) $M/M/m/K/N$: It describes a queueing system with an exponential distribution for the interarrival times of customers and the service times of customers, m servers, a maximum of K customers in the queueing system at anytime and the calling population having N potential customers.

Try an exercise.

E2) Describe the queueing systems represented by
(i) $M/D/2/N$
(ii) $GI/E_5/C/15/20/LCFS$

So far, we have discussed the physical description of the queueing system. In the following section, we shall discuss the response of these systems in terms of performance measures.

11.5 PERFORMANCE MEASURES OF QUEUEING SYSTEMS

Let us start with the customer waiting times. Customers waiting time may be considered as the time spent in the queue or the time spent in the queue plus service time. Both of the types of customers waiting time are important. If we talk about the repair of any machine, then we wish to minimize the time is queue plus service, on the other hand, if we consider the queue in an amusement park, then only waiting time in queue is important. In the same way, the length of the queue may be with or without the expected number of customers is service. Here, we are interested in determining the following basic measure of performance under steady-state conditions.

i) Expected number of customers in the queue (L_q) is the number of customers in the queue and waiting for the service. This excludes the customer being served.

$$L_q = \sum_{i=s}^{\infty} (i-s)p_i$$

where i is the number of customers in the system either waiting or in service.
s: number of parallel servers so that s customers can be served simultaneously
p_i : probability that there are exactly i customers in the queueing system under consideration in the steady state.

ii) Expected (average) number of units customers in the system, which is denoted by L_s and is given by:

$$L_s = \sum_{i=0}^{\infty} ip_i$$

or

$L_s = L_q$ + Expected number of customers in service.

iii) Expected (average) waiting time of a customer spent in the queue before being served, is denoted by W_q and is given by:

$$W_q = \frac{\text{Expected number of customers in the queue}}{\text{Arrival rate}}$$

or,

$$W_q = \frac{L_q}{\text{Arrival rate}}$$

W_q excludes the service time.

iv) Expected (average) waiting time of a customer in the system in queue and in service is denoted by W_s and is given as :

$$W_s = \frac{\text{Expected number of customers in the system}}{\text{Arrival rate}}$$

$$= \frac{L_s}{\text{Arrival rate}}$$

The formula connecting W_s and L_s is known as Little's formula ($L_s = \wedge W_s$, where $\wedge$ is the mean arrival rate.)

v) The probability that the service facility is idle or there is no customer when it is in steady state condition is p_o.

vi) Expected waiting cost per unit time is given by $= C_w L_s$, where C_w is the expected waiting cost per customer per unit time.

vii) Expected service cost per unit time $= C_s.\mu$, where C_s is the cost of servicing one unit and μ is the service rate.

viii) Total cost $C = C_w.L_s + C_s.\mu$

The total cost will be minimum if $\frac{dC}{d\mu} = 0$.

Also, the value of μ that satisfies $\frac{dC}{d\mu} = 0$ is known as minimum cost service rate.

Let us illustrate the performance measures in the following situation.

Example 2: Let us consider IGNOU PNB ATM, and observe the queue for a particular week at a particular time every day. The observations are given below:

Week days	Queue length	Service facility	Waiting time for an Arrival
1	2		3 minutes
2	4		4 minutes
3	1		2 minutes
4	0		0
5	3		1 minutes
6	4		2 minutes
7	0		0

The average length of queue $= \frac{14}{7} = 2$.

The average number of customer in the system

$$= \frac{1}{7}\left[(2+1)+(4+1)+(1+1)+(0+0)+(3+1)+(4+1)+(0+1)\right] = \frac{20}{7}.$$

The average length of non-empty queue $= \frac{2+4+1+3+4}{5} = \frac{14}{5}$.

The average waiting time of an arrival $= \frac{3+4+2+0+1+2+0}{7} = \frac{12}{7}$ min.

The average waiting time of an arrival who waits $= \frac{12}{6} = 2$ min.

Now try the following exercise.

E3) Consider a single server channel in doctor's clinic. Assume, a patient arrives and form at the end of the queue every α minutes and leaves the clinic every β minutes. Start the queue with n patients. What can you say about the queue length?

In the following section, we shall discuss the simplest single server Markovian queueing models.

11.6 THE M/M/1/∞ SYSTEM

The queueing system $M/M/1$ is also known as a simple queue or a Poisson queue or a simple Markovian.

The various assumptions about the queueing system $M/M/1$ are as follows:

i) Arrivals are described by Poisson process or equivalently, the inter-arrival time follows the exponential distribution and they come from an infinite population.

ii) Arrivals are served on a first come first served basis by a single server.

iii) Every arrival waits to be served regardless of the length of the queue.

iv) New Arrivals are independent of preceding arrivals and the average number of arrivals does not change over the time, that is the arrival rate is constant.

v) Service time also varys from one customer to the next and are independent of one another but their mean service time is same.

vi) Service times follow the exponential probability distribution or equivalently, the numbers of services are performed according to Poisson process.

The two approaches used to solve queueing models are namely mathematical and simulation. Simulation is used only if the model is mathematically intractable. In the mathematical approach, the expected values of various characteristics of queueing process are described by taking one of known mathematical distribution for arrival and service time distribution. The processes, in which arrivals and departures occur simultaneously, are known as Birth-and-Death processes. It can be recalled that this process is a special case of Markovian process, in which transitions from any state are permitted only to neighbouring states. The transition diagram of this model is given below.

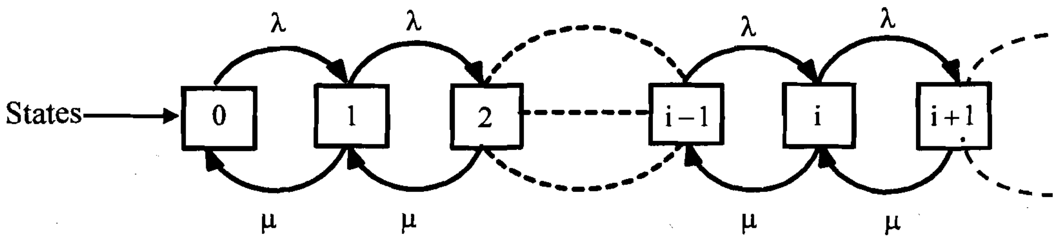


Fig. 1: Transition diagram of M/M/1

In this transition state diagram the arrival rate is denoted by λ and the service rate is denoted by μ. Let $P_i(t)$ be the probability, that the system is in state i (i customers in the queueing system) at time t, where $i = 0, 1, 2, 3, \ldots$.

Considering the transitions of the process during $(t, t+h]$, we get

$P_i(t+h) =$ the probability that the system is in the state $i+1$ and no arrival and one departure during $(t+h]+$ the probability that the system is in state i and no arrival and no departure during $(t, t+h]+$ the probability that the system is in state $i-1$ and one arrival and no departure during $(t, t+h]$.

$$= P_{i+1}(t).\,(\text{no arrival}).\,(\text{one departure})$$
$$+P_i(t).\,(\text{no arrival}).\,(\text{no departure})$$
$$+P_{i-1}(t).\,(\text{one arrival}).\,(\text{no departure})$$
$$= P_{i+1}(t).(1-\lambda h).\mu h + P_i(t)(1-\lambda h)(1-\mu h) + P_{i-1}(t)\lambda h.(1-\mu h)$$

Subtracting $P_i(t)$ both the sides and dividing by h, we get

$$\frac{P_i(t+h)-P_i(t)}{h} = P_{i+1}(t)(1-\lambda h)\mu - (\lambda+\mu)P_i(t) + \lambda\mu P_i(t)h + P_{i-1}(t)\lambda(1-\mu h)$$

Now letting $h \to 0$, we get

$$P_i'(t) = P_{i+1}(t)\mu - (\lambda+\mu)P_i(t) + \lambda P_{i-1}(t) \text{ for } i \geq 1 \qquad (1)$$

Similarly for the initial state, i.e. state 0. The probability that the system is in state 0 at time t is $P_0(t)$, which is

$$P_0(t+h) = P_0(t)(1-\lambda h) + P_1(t)\mu h$$

or,

$$\frac{P_0(t+h)-P_0(t)}{h} = -\lambda P_0(t) + \mu P_1(t)$$

after letting $h \to 0$, we get

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t) \text{ for } i = 0 \qquad (2)$$

Eqns. (1) and (2) are differential difference equations, i.e. differential equation in t and difference equation in i .The solution of Eqns.(1) and (2)will give transient state solution for the distribution of queue length.

To get the state of equilibrium or steady-state solution for P_i, which is the probability that the system is in state i at any point of time after steady state is reached as $t \to \infty$.

Then $P_i(t)$ becomes independent of time t and therefore $\lim_{t\to\infty} P_i(t) = P_i$ are constants. Consequently in the steady state we can set

$$\lim_{t\to\infty} P_i'(t) = 0 \text{ for } i = 0, 1, 2, \cdots$$

Applying these in Eqns. (1) & (2) we get

$$(\lambda+\mu)P_i = \mu P_{i+1} + \lambda P_{i-1}; \; i \geq 1 \qquad (3)$$

and $\lambda P_0 = \mu P_1$ or $P_1 = \frac{\lambda}{\mu} P_0$ (4)

Substituting $i = 1$ in eqn. (3), we get

$$(\lambda+\mu)P_1 = \mu P_2 + \lambda P_0$$

or $P_2 = \left(\frac{\lambda}{\mu}\right)^2 P_0$

Again for $i = 2$, we have

$$(\lambda+\mu)P_2 = \mu P_3 + \lambda P_1$$

or

$$P_3 = \left(\frac{\lambda}{\mu}\right)^3 P_0.$$

Continuing we can easily write

$$P_i = \left(\frac{\lambda}{\mu}\right)^i P_0. \text{ for } i = 1, 2, \ldots \qquad (5)$$

Since $\sum_{i=0}^{\infty} P_i = 1$

i.e. $\sum_{i=o}^{\infty}\left(\frac{\lambda}{\mu}\right)^{i} p_0 = 1$ [Using Eqn.(5)]. If $\frac{\lambda}{\mu}<1$, then infinite series converges and we get

$P_0 = 1-\lambda/\mu$.[for $\frac{\lambda}{\mu}>1$ the queue size will becomes explosive.]

Here, it can be seen that for P_i to be meaningful, i.e. for the existence of a steady-state solution, we should have $P_o > 0$ which happens when $\frac{\lambda}{\mu}<1$. The discrete ramdom variable $X(\text{not } X(t))$, giving the number of customers in the system in steady state is given by, $P[x=i]=P_i, i=0,1,2,\cdots$ where P_i is given by Eqn.(5), which is geometric probability distribution.

Let us try to find the various operational characteristics for this model in equilibrium state.

i) The probability that the service facility is idle i.e. there is no customer in the queue is $P_0 = 1-\lambda/\mu$.

ii) The probability that there are i customers $(i = 1, 2, 3, \ldots)$ in the queueing system is $P_i = (\lambda/\mu)^i(1-\lambda/\mu)$ (using Eqn. (5)).

iii) Expected or average number of units/customers in the queueing system is given by :

$L_s = \sum_{i=0}^{\infty} i\,P_i = \sum_{i=0}^{\infty} i(\lambda/\mu)^i(1-\lambda/\mu) = \frac{\lambda/\mu}{1-\lambda/\mu}$ or $\frac{\rho}{1-\rho}$ (in term of traffic intensity or utilisation factor ρ defined by, $\rho = \frac{\text{arrival rate}}{\text{service rate}}$, which $\frac{\lambda}{\mu}$ here)

clearly

$$L_s = \frac{\lambda}{\mu-\lambda}.$$

iv) Mean (expected or average) number of units in the queue waiting for the service, is given by :

$$L_q = \sum_{i=0}^{\infty}(i-1)P_i$$

$$= \frac{\lambda^2}{\mu(\mu-\lambda)} = \rho L_s.$$

v) Mean (expected) waiting time of a customer, in the system is:

$$W_s = \frac{L_s}{\lambda}$$

$$= 1/(\mu-\lambda) = \frac{\rho}{(1-\rho)\lambda}.$$

vi) Mean (expected or average) time a unit spends waiting in the queue:

$$W_q = \frac{L_q}{\lambda} = \frac{\lambda}{\mu(\mu-\lambda)} = \frac{\rho^2}{\lambda(1-\rho)}.$$

vii) Probability that the queue length (including the unit being served) is greater than or equal to k is

$$P(i \ge k) = \sum_{i=k}^{\infty} P_i = \sum_{i=k}^{\infty}(\lambda/\mu)^i P_0 = \frac{(\lambda/\mu)^k}{1-\lambda/\mu}.P_0$$

$$= (\lambda/\mu)^k = \rho^k$$

If the queue is non-empty, then the corresponding probability
$P(i > 1)\ 1 - P_0 - P_1$.

viii) Since the model considered here is single server model therefore the entire time of the server is alternatively a busy and a idle period. Therefore the number of idle periods would be same as that of busy period. Let the entire time be T. Then the expected duration for the server to be

Idle time $= T \times$ probability that the server is idle

$$= T.p_0 = T(1 - \lambda/\mu)$$

Expected duration for the server to be busy $= T\frac{\lambda}{\mu}$.

The expected number of idle periods $= T(1-\lambda/\mu).\lambda$.

Therefore, the mean duration of a busy period $\Rightarrow \dfrac{T.\lambda/\mu}{T\lambda(1-\lambda/\mu)}$

$$= \frac{1}{\mu - \lambda}$$

And the average number of customers served in a busy period is $\dfrac{\mu}{\mu-\lambda}$.

ix) $var(n) = E(n^2) - (E(n))^2$, where n is the number of customers in the system

$$= \sum_{i=0}^{\infty} n^2 p_n - \left(\sum_{i=0}^{\infty} np_n\right)^2$$

$$= \sum_{n=0}^{\infty} n^2.(\lambda/\mu)^n.(1-\lambda/\mu) - \left[\frac{\lambda}{\mu-\lambda}\right]^2$$

$$= \left[\frac{2(\lambda/\mu)^2}{(1-\lambda/\mu)^3} + \frac{\lambda/\mu}{(1-\lambda/\mu)^2}\right] - \left[\frac{\lambda}{\mu-\lambda}\right]^2$$

$$= \frac{\lambda}{\mu-\lambda} + \frac{\lambda^2}{(\mu-\lambda)^2} = \frac{\lambda\mu}{(\mu-\lambda)^2} = \frac{\rho}{(1-\rho)^2}.$$

x) The probability that the queue contains i customers at some time and a customer arrives during the service time of customer in service.

$$P(n = i) = \int_0^{\infty} P_i(t).\ s(t)dt$$

$$= \int_0^{\infty} \frac{(\lambda t)^i e^{-\lambda t}}{i!}.\mu e^{-\mu t} dt$$

$$= \frac{\lambda^i \mu}{i!}\int_0^{\infty} e^{-(\lambda+\mu)t} t^i dt$$

$$= \frac{\lambda^i \mu}{i!}\frac{\Gamma(n+1)}{(\lambda+\mu)^{i+1}}\left[\text{using} \int_0^{\infty} e^{-st}x^n dx = \frac{\Gamma(n+1)}{s^n}\right]$$

$$= \left(\frac{\lambda}{\lambda+\mu}\right)^i.\frac{\mu}{\lambda+\mu}[\ \text{using } \Gamma(n+1) = n!]\quad ,i = 1,2,\cdots.$$

Let us understand the above through the following examples.

Example 3: Consider the queue at a telephone booth, where only one telephone is available for making calls. The arrivals of telephone calls is according to Poisson distribution, with an average rate of 5 calls per hour and the length of telephone call is assumed to be exponentially distributed with mean 8 minutes. Find

i) the probability that the telephone is idle.

ii) the probability that there are at least two calls in the system.

iii) expected time that a call is in the queue.

Solution: Arrival rate is 5 calls per hour, i.e. $\lambda = 5$ (calls/hour)

$$\text{Service rate } = \frac{1}{8}(\text{calls/min.}) = \frac{15}{2}(\text{calls/hour})$$

$$\rho = \frac{\lambda}{\mu} = \frac{5}{15/2} = \frac{2}{3}$$

i) P (the telephone is idle) $P = (1 - \lambda/\mu) = (1 - 2/3) = \frac{1}{3}$.

ii) P(at least two calls in the system) $= P(n \geq 2) = (2/3)^2$

iii) Expected time that a call is in the queue before it is all ended

$$= W_q = \lambda/\mu(\mu - \lambda) = \frac{5}{15/2\left(\frac{15}{2} - 5\right)} = \frac{4}{15}.$$

***

Example 4: A bank has one drive in counter and all customers arrive in their cars. It is estimated that cars arrive according to Poisson distribution at the rate of 2 every 5 minutes and that there is enough space to accommodate a line of 10 cars. Other arriving cars should wait outside this space, if necessary. It takes 1.5 minutes on an average to serve a customer, but the service time actually varies according to an exponential distribution. Find

i) the probability of time the facility remains idle.

ii) the expected number of customers waiting to be served at a particular point of time.

iii) the expected time a customer spends in the system.

iv) the probability that the waiting line will exceed the capacity of the space leading to the drive.

Solution: Mean arrival rate, $\lambda = 2.60/5 = 24$ car per hr.

Mean service rate $\mu = 60/1.5 = 40$ cars per hr.

i) The proportion of time, the facility remains idle:
$P_0 = 1 - \lambda/\mu = 1 - 24/20 = 0.4$.

ii) The expected number of customers in the waiting line:
$L_q = \lambda^2/\mu.(\mu - \lambda) = 24.24/40(40 - 24) = 576/64 = 0.9$.

iii) The expected time of customer in the system:
$W_s = 1/(\mu - \lambda) = 1/40 - 24 = 1/16$ hrs $= 3.75$ minutes.

iv) Waiting time will exceed their capacity of the space if 11 or more cars are arrived there because the space can accommodate only 10 cars. Hence the required probability can be worked out as under:
$P(n \geq 11) = (\lambda/\mu)^{11} = (24/40)^{11} = 0.0036$.

***

Example 5: The mean rate of arrival of planes at an airport during the peak period is 20 per hour and the number of arrivals follow a Poisson process. The airport can land 60 planes per hour on an average in good weather and 30 planes per hour in bad weather, but the number landed follow a Poisson process with these respective rates. When there is congestion, the planes that arrived earlier,

i) How many planes would be flying over the field in the stack on an average in good weather and in bad weather?

ii) How long a plane would be in the stack and in the process of landing in good weather, and in bad weather?

iii) How long a plane would be in the process of landing in good weather and in bad weather after stack awaiting?

Solution:

Arrival rate (λ) in good weather = 20 planes/hour

Arrival rate (λ) in bad weather = 20 planes/hour

Service rate (μ_1) in good weather = 60 planes/hour

Service rate (μ_2) in bad weather = 30 planes/hour

i) Average number of the planes in queue in good weather

$$= \frac{\lambda^2}{\mu_1(\mu_1 - \lambda)} = \frac{400}{60 \times 40} = \frac{1}{6}$$

Average number of the planes in queue in bad weather $= \frac{400}{30 \times 10} = \frac{4}{3}$

ii) Average waiting time in the system in good weather

$$= \frac{1}{\mu_1 - \lambda} = \frac{1}{60 - 20} = \frac{1}{40} \text{hour} = 1.5 \text{min}$$

Average waiting time in the system in bad weather $= \frac{1}{30 - 20} = \frac{1}{10} \text{hour} = 6 \text{min}$

iii) Average service time in good weather is $\frac{1}{\mu_1} \times 60 = 1$ minute

i.e. (Average waiting time in system - Average waiting time in stack)= (1.5min – 0.5)

Average service time in bad weather is 2 minutes

i.e. (Average waiting time in system - Average waiting time in queue) = (6 min – 4 min)

***

You may try the following exercises.

E4) To support National Heart Week, the Heart Association plans to install a free blood pressure testing booth in Ansals Mall for the week. Previous experience indicates that, on the average, 10 persons per hour request a test. Assume arrivals are Poisson from an infinite population. Blood pressure measurements can be made at a constant time of five minutes each. Determine what is

a) the average number of persons in line will be.

b) the average number of persons in the system.

c) the average amount of time a person can expect to spend in line before his blood pressure in checked.

d) on an average, the time taken to measure a person's blood pressure, including waiting time.

Additionally, on weekends, the arrival rate can be expected to increase to nearly 12 per hour.

What effect will this have on the number in the waiting line?

E5) A childcare shop dealing with children's requirements, has one cashier who handles all customers' payments. The cashier takes on an average 4 minutes per customer. Customers come to cashiers area in a random manner but on an average rate of 10 per hour. The management received a large number of customer's complaints and decided to investigate the following questions. Answer these questions.

a) What is the average length of the waiting line to be expected under the existing conditions?

b) What portion of time is the cashier expected to be idle?

c) What is the average length that a customer would be expected to wait to pay for his purchase?

d) If it was felt that a customer would not tolerate a wait of more than 12 minutes, what is the probability that a customer would have to wait at least this length of the time?

E6) A milk plant distributes its product by trucks, loaded at the loading dock. It has its own fleet plus the trucks of a transport company are used. This company has complained that sometimes the trucks have to wait in the queue and thus the company loses money. The company has asked the management either to go in for a second loading dock or discount prices equivalent to waiting time. The data available is as follows:

Average arrival rate = 3per hour, and average service rate = 4 per hour.

The transport company has provided 40% of the total number of trucks. Stating assumptions determine:

a) The probability that a truck has to wait.

b) The average waiting time of a truck.

c) Expected waiting time for the transport company trucks per day.

E7) Workers come to tool storeroom to enquire about special tools (required by them) for accomplishing a particular project assigned to them. The average time is two arrivals per 50 seconds and the arrivals are assumed follow in Poisson process. The average time to attend enquiry (by the tool room attendant) is 30 seconds. Determine:

i) average queue length.

ii) average length of non empty queue.

iii) average number of workers in the system including the worker being attended.

iv) mean waiting time of an arrival.

v) average waiting time of an arrival who waits

vi) the type of policy to be established.

In other words, determine the additional number of storeroom attendants, which will minimize the combined cost of attendant's idle time and the cost of workers' waiting time. Assume the charges of a skilled worker Rs.4 per hour and that of tool room attendant Re. 0.75 per hour.

Above we discussed M/M/1 queueing process as a simple birth-death process with constant birth-death rates. Let us now model a queueing process when the arrival and service rates are state dependent. Again basic differencedifferential equation follows from a birth-death process.

The transition diagram of this system is given below

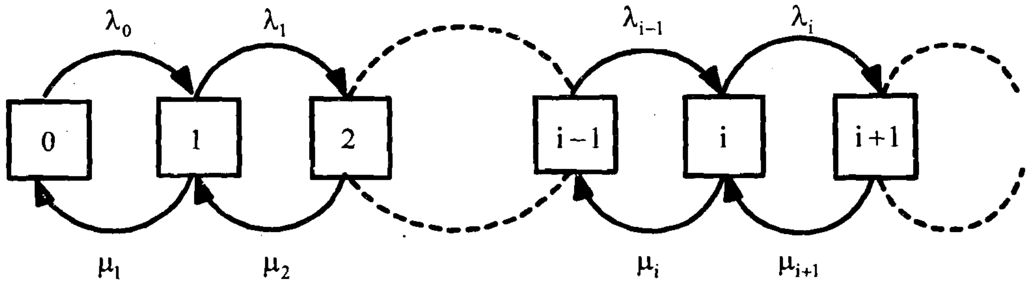


Fig. 2: Transition diagram

The assumptions are given below:

λ_i : state dependent arrival rate when i customers are in the system.

μ_i : state dependent departure rate when i customers are in the system.

i : number of customers in the system (in queue and in service).

P_i : steady state probability of the i^{th} state, i.e. i customers are in the system.

The difference-differential equations of the model now modify to

$$P_i'(t) = -(\lambda_i + \mu_i)P_i(t) + \lambda_{i-1}P_{i-1}(t) + \mu_{i+1}P_{i+1}(t),\ i \geq 1 \qquad (6)$$

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t)\, P \qquad (7)$$

Applying the steady state conditions, we get

$$P_{i+1} = \frac{\lambda_i + \mu_i}{\mu_{i+1}} P_i - \frac{\lambda_{i-1}}{\mu_{i+1}} P_{i-1},\ i \geq 1 \qquad (8)$$

and $$P_1 = \frac{\lambda_0}{\mu_1} P_0. \qquad (9)$$

Substituting $i = 1$ in Eqn. (8) and using Eqn. (9), we get

$$P_2 = \frac{\lambda_1 \lambda_0}{\mu_1 \mu_2}.$$

Continuing or by applying mathematical induction, we get

$$P_i = \frac{\lambda_0 \lambda_1 \ldots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} P_0 \quad ,i = 1, 2, \cdots \qquad (10)$$

Now as $\sum_{i=0}^{\infty} P_i = 1$,

it gives $$P_0\left[1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\ldots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i} + \cdots\right] = 1, \qquad (11)$$

Various operational characteristics can easily be written down from steady state solution using above.

Now try to solve another queueing model in the following exercise.

E8) Consider an $M/M/1$ queueing system where arrival rate is λ and service rate is μ per unit time, where $\lambda < \mu$. Find the equilibrium probability P_n if not more than N customers are allowed in the queueing system.

So far we have discussed the queueing systems with one server. In the next section, we shall focus on the models having multiple-servers.

11.7 M/M/k /∞ SYSTEM

In a multiple-server queueing system, two or more servers (or channels) are available to handle customers who arrive for service. It covers the situation, where for example, there may be more than one runway at an airport for takeoff and landing; there may be more than one doctor in a hospital O.P.D whom the patients can visit; there may be more than one teller in a bank, and so on.

The most common and basic multiple service channel system contains parallel stations serving a single queue on a first come first served basis.

The Assumptions for this queueing system are as follows:

1) The input population is infinite.
2) A single waiting line is formed.
3) The service is on a first come first served basis.
4) The arrival of customers follows Poisson probability law and service time has an exponential distribution.
5) There are k service stations, each of which provides identical service.
6) The arrival rate is smaller than the combined service rate of all k service facilities.

With the number of calling units represented by n and the number of channels or service stations by k, the descriptive characteristics of the multi channel system are summarized as follows:

$\lambda =$ average rate of arrivals
$\mu =$ average service rate of each of the channel (or servers)
$k\mu =$ mean combined service rate of all the channels (or servers)
$\rho = \lambda / k\mu =$ utilization factor of the entire system

Also, in terms of the generalized birth-death model the rates can be expressed as

$$\lambda_n = \lambda \quad ; i \geq 0$$

$$\mu_n = \begin{cases} i\mu \; ; \; i < k \\ k\mu \; ; \; i \geq k \end{cases}$$

To solve this queueing model, let us first write the difference differential equations. Using Eqns. (6) and (7) as follows

$$P_i'(t) = -(\lambda + i\mu)\, P_i(t) + (i+1)\mu P_{i+1}(t) + \lambda P_{i-1}(t);\ 1 \leq i < k \tag{12}$$

$$P_i'(t) = -(\lambda + k\mu) P_i(t) + k\mu P_{i+1}(t) + \lambda P_{i-1}(t);\ i \geq k \tag{13}$$

and

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t) \tag{14}$$

The steady state equations from Eqn. (8) are

$$(i+1)\mu P_{i+1} = (\lambda + i\mu)P_1 - \lambda P_{i-1}; 1 \le i < k \tag{15}$$

$$k\mu P_{i+1} = (\lambda + k\mu)P_i - \lambda P_{i-1};\ i \ge k \tag{16}$$

and $$\lambda P_0 = \mu P_1 \text{ or } P_1 = \frac{\lambda}{\mu}P_0 \tag{17}$$

Substituting $i = 1, 2, \ldots$ in Eqn.(15), we have

$$2\mu P_2 = (\lambda + \mu)P_1 - \lambda P_0$$

$$P_2 = \frac{\lambda+\mu}{2\mu}.\frac{\lambda}{\mu}.P_0 - \frac{\lambda}{2\mu}P_0 \text{ [using Eqn.(17)]} \tag{18}$$

$$= \left(\frac{\lambda+\mu}{2\mu}.\frac{\lambda}{\mu} - \frac{\lambda}{2\mu}\right)P_0 = \frac{\lambda}{\mu}.\frac{\lambda}{2\mu}P_0 = \frac{\lambda^2}{2!\mu^2}P_0$$

For $i = 2$

$$3\mu P_3 = (\lambda + 2\mu)P_2 - \lambda P_1$$

$$P_3 = \frac{\lambda+2\mu}{3\mu}P_2 - \frac{\lambda}{3\mu}P_1$$

$$= \frac{\lambda+2\mu}{3\mu}.\frac{\lambda^2}{2\mu^2}P_0 - \frac{\lambda}{3\mu}.\frac{\lambda}{\mu}P_0 \quad \text{[using (17) and (18)]}$$

$$= \frac{\lambda}{\mu}.\frac{\lambda}{2\mu}.\frac{\lambda}{3\mu}P_0 = \frac{\lambda^3}{3!\mu^3}P_0$$

Continuing in the manner or by using the mathematical induction, we can write

$$P_i = \frac{\lambda^i}{i!\mu^i}P_0;\ 1 \le i < k \tag{19}$$

Substituting $i = k$ in Eqn. (16), we get

$$k\mu P_{k+1} = (\lambda + k\mu)P_k - \lambda P_{k-1}$$

$$P_{k+1} = \frac{\lambda + k}{k\mu}P_k = \frac{\lambda}{k\mu}P_{k-1}$$

$$= \frac{\lambda+k\mu}{k\mu}\left[\frac{\lambda+\mu}{k\mu}P_{k-1} - \frac{\lambda P_{k-2}}{k\mu}\right] - \frac{\lambda}{k-\mu}P_{k-1} \quad \text{[using Eqn.(15)]}$$

$$= \left(\frac{(\lambda+k\mu)\,\{\lambda+(k-1)\mu\}}{(k\mu)\,(k\mu)} - \frac{\lambda}{k\lambda}\right)\left(\frac{\lambda^{k-1}}{(k-1)!\mu^{k-1}}\right)P_0 - \frac{\lambda(\lambda+k\mu)}{(k\mu)\,(k\mu)}\frac{\lambda^{k-2}}{(k-2)!\mu^{k-2}}P_0$$

[using Eqn. (19)]

or $$P_{k+1} = \frac{\lambda^2}{k\mu}.\frac{\lambda^{k-1}}{k!\mu^k}P_0$$

$$= \frac{\lambda^{k+1}}{k.k!\mu^{k+1}}P_0.$$

Similarly

$$P_{k+2} = \frac{\lambda^{k+2}}{k^2k!\mu^{k+2}}P_0$$

In general, we can write

$$P_i = \frac{\lambda^i}{k^{i-k}k!\mu^k}P_0;\ i \ge k \tag{20}$$

Now let us find the operational characteristic of this model.

i) **Probability that the system shall be idle**

For this we must find P_0, which can be obtained from Eqn. (11) as show below.

We always have

$$P_0 + P_1 + P_2 + \cdots + P_{k-1} + P_k + P_{k+1} + \cdots = 1$$

or $$P_0\left[\sum_{i=1}^{k-1}\frac{\lambda^i}{i!\mu^i} + \sum_{i=k}^{\infty}\frac{\lambda^i}{k^{i-k}k!\mu^i}\right] = 1$$

or $$P_0\left[\sum_{i=1}^{k-1}\frac{\lambda^i}{i!\mu^i} + \frac{k^k}{k!}\sum_{i=k}^{\infty}\left(\frac{\lambda}{k\mu}\right)^i\right] = 1$$

or $$P_0\left[\sum_{i=1}^{k-1}\frac{\lambda^i}{i!\mu^i} + \frac{k^k}{k!}\cdot\frac{(\lambda/k\mu)^k}{1-\dfrac{\lambda}{k\mu}}\right] = 1 \text{ [as } \lambda < k\mu]$$

or $$P_0 = \left[\sum_{i=1}^{k-1}\frac{\lambda^i}{i!\mu^i} + \frac{k\mu.\lambda^k}{k!(k\mu-\lambda)\mu^k}\right]^{-1} \qquad (21)$$

ii) **Probability that there shall exactly be n customers in the system:**

$$P_n = \begin{cases} \dfrac{\lambda^n}{n!\mu^n}.P_0\,; & 1 \le n < k \\ \dfrac{\lambda^n}{k^{n-k}k!\mu^k}.P_0; & n \ge k \end{cases}$$

iii) **The expected or average number of customers (or units) in the waiting line:**

$$L_q = \sum_{i=k}^{\infty}(i-k)P_i = \frac{\lambda\mu(\lambda/\mu)^k P_0}{(k-1)!(k\mu-\lambda)^2}$$

iv) **The expected number of customers being served or the average number of busy servers:**

$$L_s = \sum_{i=0}^{k-1} i\, P_i + \sum_{i=k}^{\infty} k\, P_i = \lambda/\mu$$

v) **The average number of customers in the system is**

$$L = L_s + L_q$$

vi) **The average time a customer (or unit) spends in the queue waiting for service:**

$$W_q = L_q/\lambda$$

vii) **The average time a unit spends in the waiting line and being serviced (namely in the system):**

$$W_s = W_q + 1/\mu.$$

Let us illustrate the following examples.

Example 6: Plans are being made for a plant enlargement. Repair facilities for machine breakdowns are barely adequate in the existing plant and will certainly not provide acceptable service when more machines are added. Records of recent repair activities show an average of four breakdowns per 8-hour shift. The pattern of breakdowns closely follows a Poisson process. When the new additions to the plant are completed, an average of six breakdowns per shift following the same process is expected. An exponentially distributed service at a rate of six repairs per shift is the capacity of the present repair facility.

Two alternatives with equivalent annual cost are available. New equipment and a larger crew for the existing station would increase the average servicing rate to 11 repairs per shift or a second servicing station could be built in the new addition. In the latter alternative, the capacity of the two service stations would be five servicing per shift in each. Repair times would be exponentially distributed.

Which of the two alternatives would be more efficient in terms of customer waiting time?

Solution: The single-channel alternative would have the following characteristics when

$\lambda = 6$ and $\mu = 11$

Probability of being idle: $P_0 = 1 - \lambda/\mu = 1 - 6/11 = 0.45$

Expected number of machines waiting for service:

$L_q = \lambda^2/\mu(\mu - \lambda) = 6^2/11(11-6) = 0.66$.

Expected time before a machine is repaired:

$W_q = 1/\mu - \lambda = 1/11 - 6$

$= 0.2 \times 8$ hr./shift $= 1.6$ hr.

The values of the same characteristics for the second alternatives providing two channels, $k = 2$ each with $\mu = 5$ and $\lambda = 6$ are:

$$P_0 = \left[\sum_{n=0}^{k-1} (\lambda/\mu)^n / .n! + (\lambda/\mu)^k / k!(1 - \lambda/k\mu)\right]^{-1}$$

$$= \left[\sum_{n=0}^{2-1} (6/5)^n / n! + (6/5)^2 / 2!\{1 - (6/5 \times 2)\}\right]^{-1}$$

$$= [1 + 6/5 + (36/25)/2\ (4/10)]^{-1}$$

$$= 0.25$$

$$Lq = \{\lambda\mu(\lambda/\mu)^k / (k-1)!(k\mu - \lambda)^2\}$$

$$= \{6 \times 5 \times (6/5)^2 / (2-1)!(2 \times 5 - 6)^2\}\ (0.25) = 0.68 \text{ machine.}$$

$$W_q = L_q/\lambda + 1/\mu = 0.68/6 + 1/5 = 0.31 \text{ shift}$$

$$= 0.31 \times 8 \text{ per hr./shift} = 2.5 \text{ hr.}$$

***

First alternative smaller L_q and W_q and hence is a better alternative.

Example 7: Bank customers can be served from any one of three different service points. They are observed to enter the bank in a Poisson fashion at an average rate of 24 per hour, whilst the average time a bank clerk takes to service a customer is considered to be exponentially distributed at an average time of 6 minutes.

As an alternative, the bank is thinking of installing an automatic servicing machine which although means only one service channel, but it will be able to serve individual customers three times as the bank clerks do at present.

Advise the bank on which of the systems would be the more efficient in terms of customer waiting time.

Solution: Current system: here $\lambda = 24$, $\mu = (60/6) = 10$ and $k =$ no. of channels $= 3$

The expected waiting time in queue:

$$W_q = \{\mu(\lambda/\mu)^k /(k-1)!\,(k\mu-\lambda)^2\}.P_0 \text{ ,where}$$

$$P_0 = \left[\sum_{n=0}^{k-1}(\lambda/\mu)^n/n! + (\lambda/\mu)^k/k!(1-\lambda/k\mu)\right]^{-1}$$

$$W_q = \{\mu(\lambda/\mu)^3/(3-1)!\,(3\mu-\lambda)^2\}.\,[1+\lambda/\mu+(1/2!).\,(\lambda/\mu)^2+(1/3!).(\lambda/\mu)^3.(3\mu/3\mu-\lambda)]^{-1}$$

$$=\{(24/10)^3 10/2!(3\times 1024)\}.[1+24/10+1/2.\,(24/10)^2+1/6(24/10)^3.(3\times 10/3\times 10-24)$$

$$=138.24/72[1+2.4+2.88+11.52]^{-1}$$

$$=138.24/72\times 17.8 = 0.1079 \text{ hour, or } 6.5 \text{ minute.}$$

Proposed System: here $\lambda = 24$ per hr. and $\mu = 30$ per hr.

Average time in queue;

$$W_q = \lambda/\mu.(\mu-\lambda) = 24/30\,(30-24) = 4/30 \text{ hr. or 8 minutes.}$$

Hence in terms of waiting time, the current multi-channel system is superior.

***

Try the following exercises.

E9) A general insurance company has three claims adjusters in its branch office. People with claims submitted against the company are found to follow a Poisson process, at an average rate of 20 per- 8 hour day. The amount of the time that an adjuster spends with a claimant is found to have a negative exponential distribution, with mean service time 40 minutes. Claimants are processed in the order of their appearance:

a) How many hours a week an adjuster is expected to spend with claimants?

b) How much time on the average does a claimant spend in the branch office?

E10) A post office has two counters, the first counter handles money orders and registration letters only and the second handles all other businesses. It has been found that service time distribution for both the counters are exponential with mean service time of 3 minutes per customer. The customers found to arrive at the first counter in a Poisson fashion with mean arrival rate of 14 per hour. While at the second counter at a rate of 16 per hour. What should be the effect on average waiting time for two type of customers if each counter can handle all the types of business? What should be the effect if this could only be accomplished by increasing the mean service time to 3.5 minutes?

E11) The management of the Alexandria Transport Service Company Ltd. is concerned about the duration of the time the company's trucks are idle, waiting to be unloaded. The terminal operates with four unloading bays. Each bay requires a crew of two employees, and each crew costs Rs 50 per hour. The estimated cost of an idle truck is Rs 200 per hour. Trucks arrive at an average rate of three per hour, according to Poisson process. A crew can unload a semitrailer rig in an average time of one hour, with exponential service times. What is the total hourly cost of operating the system?

Let us now summarize this unit.

11.8 SUMMARY

In this unit, we have discussed

i) Queues are the waiting lines which we observe at many places in daytoday life.

ii) Characteristics of a queue like arrival pattern, service mechanism, servers, queue discipline, customer's behaviour, size of population and length of the queue.

iii) Degenerate distribution, negative exponential distribution and general distributions, which are applied for arrival and service patterns.

iv) Performance measures are computed under steady-state conditions.

v) Single server model $M/M/1/\infty$ and $M/M/k/\infty$ in details with operational characteristics.

11.9 SOLUTIONS/ANSWERS

E2) i) $M/D/2/N$ would describe as given below
M : Poisson arrival process
D : Deterministic service times
2 : Two servers
N : System capacity of N.

ii) GI : General independent interarrival time
E5 : Erlang-5 service time
C : C servers
15: System capacity of 15
20: A finite source of size 20
LCFS : Last came-first served queue discipline.

E3) Given that

$$\text{arrival rate} = \frac{1}{\alpha} \text{ per min.}$$

$$\text{depature rate} = \frac{1}{\beta} \text{ per min.}$$

there may arise three cases

i) If $\alpha > \beta$, then the patients arrive faster than they leave. In this case, the queue will grow infinitely large.

ii) If $\alpha = \beta$, then the queue length will remain n always.

iii) If $\alpha < \beta$, then the queue length will follow n and the queue may end. In this case, the time for which doctor is idle may also be of interest.

E4) From the given data, we have:
Mean rate of arrival, $\lambda = 10$ patient/hour

Mean rate of service, $\mu = \frac{60}{5} = 12$ patient/hour

a) Mean (expected) number of units in the queue waiting for service:

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)} = \frac{102}{12(12-10)} = \frac{100}{12\times 2} = \frac{25}{6}.$$

b) Average no. of persons in the system:

$$L_s = \frac{\lambda}{\mu-\lambda} = \frac{10}{12-10} = 5.$$

c) Average amount of time a person can expect to spend in line

$$= \frac{\lambda}{\mu(\mu-\lambda)} = \frac{10}{12(12-10)} = \frac{5}{12} \text{ hours.}$$

d) Average time will take to measure a person's blood pressure, including waiting time $= \frac{1}{\mu - \lambda} = \frac{1}{12}$ hour.

e) Now the average rate of arrival $\lambda = 12/\text{hr}$. Average no. of people in waiting line $= \frac{\lambda^2}{\mu(\mu - \lambda)} = 0$.

E5) From the given set of data
Mean rate of arrival, $\lambda = 10$ customers/hour
Mean rate of service, $\mu = 15$ customers/hour

a) Average number of customers in the queue:
$L_q = \lambda^2 / \mu\,(\mu - \lambda) = 10.\,10/15\,(15-10) = 4/3$.

b) Probability that the cashier will be idle:
$P_0 = 1 - (\lambda/\mu) = 1 - 10/15 = 1/3$
It implies that the cashier will remain idle 33.3% of his time.

c) Average length of time that a customer is expected to wait in a queue:
$W_q = \lambda/\mu\,(\mu - \lambda) = 10/15\,(15-10)$ hrs or 8 minutes.

d) Probability that a customer has to wait at least 12 minutes before he can make payment $=$ P(waiting time $\geq 1/5$ hr.) $= (\lambda/\mu)e^{-(\lambda-\mu)t} = 2/3\, e^{-(10-15)/5} = 2/3(0.368) = 0.245$.

E6) From the given data, it is observed
$\lambda = 3$ trucks per hour, $\mu = 4$ per hour

i) Probability that a truck has to wait for service
$= 1 - P_0 = 1 - (1 - \lambda/\mu) = \lambda/\mu = 3/4 = 0.75$

ii) Expected waiting time of a truck $= W_s = 1/(\mu - \lambda) = 1/4 - 3$ or 1 hour.

iii) Expected waiting time of a company's truck per day = (number of trucks per day) × (% contractors trucks) × (Expected waiting time per truck)
$= 3.24.(40/100).(3/4(43)) = 21.6$ hrs.

E8) The steady state difference differential equations for this model are
$\lambda P_0 = \mu P_1$; $i = 0$
$\mu P_{i+1} = (\lambda + \mu)P_i + \lambda P_{i-1}$; $i = 1, 2, \ldots, N-1$
$\lambda P_{N-1} = \mu P_N$; $i = N$
After solving the equation above, we get

$$P_0 = \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{N+1}}; \quad \lambda < \mu$$

$$P_i = \begin{cases} \frac{1 - \lambda/\mu}{1 - (\lambda/\mu)^{N+1}}; & \lambda < \mu; 1 \leq N; \lambda < \mu \\ \frac{1}{N+1} & ; \lambda = \mu \end{cases}$$

E9) For the given data
$\lambda = \frac{20}{8} = \frac{5}{2}$ / day and $\mu = \frac{60}{40}$ / day $= \frac{3}{2}$
there are three adjuster in the system, so
$k = 3$.

Thus, $P = \frac{\lambda}{k\mu} = \frac{5}{9}$

Probability that no customer in the system

$$P_0 = \left[\sum_{n=0}^{k-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \left(\frac{k\mu}{k\mu - \lambda}\right)\right]^{-1}$$

$$= \left[\sum_{n=0}^{2} \frac{1}{n!}\left(\frac{5}{3}\right)^n + \frac{1}{3!}\left(\frac{5}{3}\right)^3 \left(\frac{3\times 3/2}{3\times 3/2 - 5/2}\right)\right]^{-1}$$

$$P_0 = \frac{24}{139}$$

and $P_n = \frac{1}{k!\,k^{n-k}}\left(\frac{\lambda}{\mu}\right)^n P_0$.

Thus, $P_1 = \frac{40}{139}$ and $P_2 = \frac{100}{147}$.

Expected number of adjusters at any specified time will be

$3P_0 + 2P_1 + P_2 = 4/3$.

Probability (one adjuster is idle) $= \frac{4}{3\times 3} = \frac{4}{9}$

Probability (adjuster is busy) $= 1 - \frac{4}{9} = \frac{5}{9}$.

a) No. of hours a week an adjuster has to spend with claimants

$= \frac{5}{9}\times 8\times 5 = 22.2$ hrs.

b) No. of hrs. a claimant has to spend in a branch office

$$W_s = \frac{L_q}{\lambda} + \frac{1}{\mu}$$

$$= \frac{1}{k-1!}\left(\frac{\lambda}{\mu}\right)^k \cdot \frac{\lambda\mu}{(k\mu-\lambda)^2} P_0 + \frac{1}{\mu}$$

$W_s = 49$ minutes.

E10) Initially there are two independent queues, where one is for money orders and registration letters and second is for other work. From the above set of data:

For the first counter:

$\lambda = 14/\text{hr}.\quad \mu = 3/\text{minutes or } 20/\text{hr}.$

Average time in queue $W_q = \lambda/\mu(\mu-\lambda) = 14/20\ (20-14) = 7/60$ hrs or 7 minutes.

For the second counter:

$\lambda = 16/\text{hr}.\quad \mu = 3 \text{ minutes or } 20/\text{hr}.$

Average time in queue $W_q = \lambda/\mu\ (\mu-\lambda) = 16/20\ (20-16) = 1/5$ hrs or 12 minutes.

If each server can handle both the counters, then we can have a single queue for both types of services.

The queuing system is thus with 2 service channels with $\lambda = 14 + 16 = 30/\text{hrs}$ and $\mu = 20/\text{hrs}$ and $k = 2$, $P = \frac{\lambda}{k\mu} = \frac{3}{4}$.

$$P_0 = \left[\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{k!}\left(\frac{\lambda}{\mu}\right)^k \left(\frac{k\mu}{k\mu-\lambda}\right)\right]^{-1}$$

$$= \left[\sum_{n=0}^{1} \frac{1}{n!}\left(\frac{3}{2}\right)^n + \frac{1}{2!}\left(\frac{3}{2}\right)^2 \left(\frac{40}{40-30}\right)\right]^{-1}$$

$$= \left[1 + \frac{3}{2} + \frac{1}{2}.\frac{9}{4}.4\right]^{-1} = \frac{1}{7}.$$

Average waiting time of arrivals in the queue.

$$W_q = \frac{L_q}{\lambda} = \left[\frac{1}{(k-1)!}\left(\frac{\lambda}{\mu}\right)^k \frac{\mu}{(k\mu-\lambda)^2}\right] P_0$$

$$= \left(\frac{3}{2}\right)^2 . \frac{20}{(40-30)^2} . \frac{1}{7} = \frac{9}{140} \text{ hr. or 3.83 minutes.}$$

Combined waiting time with increased service time, when

$\lambda = 30/\text{hr.}, \mu = 60/3.5$ or $\frac{120}{7}/\text{hr.}$

$$P_0 = \left[\sum_{n=0}^{1} \frac{.1}{n!}\left(\frac{21}{12}\right) + \frac{1}{2!}\left(\frac{21}{12}\right)^2 \frac{.2.(120/7)}{2.(120/7)-30}\right]^{-1}$$

$$= \left[\frac{7}{4} + \frac{1}{2}.\frac{49}{16}.\frac{240}{30}\right]^{-1} = \frac{1}{15}.$$

Average waiting time of arrivals in the queue

$$W_q = \frac{1}{(k-1)!}\left(\frac{\lambda}{\mu}\right)^k \frac{\mu}{(k\mu-\lambda)^2} . P_0$$

$$= \left(\frac{7}{4}\right)^2 . \frac{120/7}{(2.120/7-30)^2} . P_0$$

$$= 11.43 \text{ minutes.}$$

E11) It is the case of multiple servers. To find the total cost of labour and idle trucks, we must calculate the average waiting time in the system. However, we need to calculate the average number of trucks in queues and the average waiting time in queue.

The average utilization of the four bays is:

$$\rho = \lambda / k\mu = 3/4 \times 1 = 0.75 \text{ or } 75\%$$

for the level of utilizations, we can now compute the probability of no trucks in the system as follows:

$$P_0 = \left[\sum_{n=0}^{k-1} \left(\frac{\lambda}{\mu}\right)^n / n! + \left(\frac{\lambda}{\mu}\right)^k / k!\ (1 - \lambda / k\mu)\right]^{-1}$$

$$= [1 + 3 + 9/2 + 27/6 + 81/24\ (1-0.75)]^{-1}$$

$$= 0.0377.$$

The average number of trucks in queues:

$$L_q = [(\lambda/\mu)^k \rho / k!\ (1-\rho)^2].\ P_0$$

$$= [(3/1)^4\ (0.75)/4!(1-0.75)^2] \times (0.0377) = 1.53 \text{ trucks}$$

The average waiting time in queues:

$$W_q = L_q / \lambda = 1.53/3 = 0.51 \text{ hours}.$$

The average waiting time spent in the system:

$$W_s = W_q + 1/\mu = 0.51 + 1 = 1.51 \text{ hours}.$$

The average number of trucks in the system is:

$$L_s = \lambda\ W_s = 3\ (1.51) = 4.53 \text{ trucks}.$$

Thus, the number of trucks in the system averages 4.53 at all times. So we can calculate the hourly cost of labour and idle trucks as follows:

Labour cost $= 50\ k = 50.4 =$ Rs. 200

Idle truck cost $=$ Rs. 200 $L_s =$ Rs. 200(4.53) $=$ Rs. 906

Total hourly cost $=$ Rs.1,106.00 .

—x—